Multilayer Perceptron Network of ECG Peaks for Cardiac Abnormality Detection

Mohd Salman Mohd Sabri
Faculty of Engineering
Universiti Pertahanan Nasional Malaysia, Sg. Besi Camp
Kuala Lumpur, Malaysia
salman@upnm.edu.my

Nur Izzani Mat Rozi
Faculty of Engineering
Universiti Pertahanan Nasional Malaysia, Sg. Besi Camp
Kuala Lumpur, Malaysia
izzani299@gmail.com

Shazreen Shaharuddin
Faculty of Medical & Defense Health
Universiti Pertahanan Nasional Malaysia, Sg. Besi Camp
Kuala Lumpur, Malaysia
shazreen@upnm.edu.my

Maizatullifah Miskan
Faculty of Medical & Defense Health
Universiti Pertahanan Nasional Malaysia, Sg. Besi Camp
Kuala Lumpur, Malaysia
maizatullifah@upnm.edu.my

Fakroul Ridzuan Hashim
Faculty of Engineering
Universiti Pertahanan Nasional Malaysia, Sg. Besi Camp
Kuala Lumpur, Malaysia
fakroul@upnm.edu.my

Mohd Sharil Saleh
Faculty of Electrical & Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat
Johor, Malaysia
sharil@upnm.edu.my

***Abstract*—The inception of artificial neural networks (ANNs) was predicated on computational adaptations of human biology, specifically the fundamental principles underlying neurons. The feasibility of utilising ANN for diverse problem domains has been extensively investigated, with a particular emphasis on the domain of biomedical engineering. Applications of ANN are commonly employed in the fields of medicine and education for decision-making purposes. The ANNs employed in the present investigation were trained to identify cardiac anomalies by utilising a diverse set of reference data. The input parameters utilised for cardiac difficulties are commonly known as reference parameters, specifically pertaining to the amplitude and duration of the electrocardiogram (ECG) signal. The ECG complex is composed of three distinct components: the P peak, the QRS wave, and the T peak. The artificial neural network is provided with six input parameters, which are obtained by measuring the amplitude and length of each P peak, QRS wave, and T peak. The present study utilises a multilayer perceptron (MLP) as the structure for the ANN. This study examines the impact of the Tansig and Purelin activation functions on the structure of the MLP. All other networks were not as good as the MLP network, which got the best performance of 96.32% by using the BayR training method and the Tansig activation function.**

***Keywords*—ECG, amplitude, MLP network, activation function**

I. Introduction

Arrhythmia refers to a specific type of cardiac irregularity or abnormality in heart rhythm that commonly manifests as unexpected heartbeats in individuals. Several aberrations include activity-triggered, re-entry, and induced fibrillation. Cardiac rhythm abnormalities encompass a diverse range of pathological conditions characterised by aberrant electrical activity inside the heart. Aberrant cardiac rhythms can manifest as tachycardia, bradycardia, regularity, irregularity, or a mix thereof [1-2]. In Malaysia, depression has emerged as the second leading cause of mortality, following cardiovascular disease, which has demonstrated an increasing trend. If prompt monitoring and/or early diagnosis are implemented, there is potential for a decrease in the prevalence of this illness. Arrhythmia, a cardiac phenomenon that can serve as a diagnostic tool for early detection of problems, provides clinicians with valuable insights to determine the optimal course of treatment [1-3]. The utilisation of ECG is crucial for the monitoring of episodes of myocardial infarction, also known as heart attacks, with subsequent emphasis placed on preventive measures. The ECG signal can be analysed by examining the P peak, QRS wave, and T peak. The occurrence of a P peak is attributed to atrial depolarization, while a QRS complex is indicative of ventricular depolarization, and a T peak signifies ventricular repolarization. The observed peaks in the data may potentially be attributed to cardiac electrical activity [1-2].

The analysis of ECG data is commonly used in clinical studies as a method for detecting cardiac activity. The heart activity detection industry commenced in the latter half of the 1950s. According to statistical data, there is an increasing number of recordings being assessed on an annual basis. Consequently, there has been a surge in research pertaining to the automated processing of ECG signals, which has subsequently given rise to the establishment of the field of biomedical engineering as an engineering discipline. In the preceding two to three decades, a multitude of methods for analysing ECG signals have been developed and implemented. The detection of the QRS complex is the primary focus of numerous investigations. Several parameters in the ECG signal recording can provide insights into the patient's physiological condition [4-6]. Applications such as ANNs are modified in accordance with empirical models that accurately represent the functioning of the human brain [7]. ANNs possess capabilities such as the ability to map input data to output data, solve both linear and non-linear problems, and exhibit neurobiological similarities. ANNs have found extensive use in several domains such as research, technology, finance, and education, particularly in larger-scale applications. ANN are widely employed in the field of engineering for various purposes, including but not limited to pattern recognition, data categorization, system identification, image processing, and accuracy enhancement.

This study aims to investigate the efficacy of forecasts in accurately identifying heart failure. The prediction procedure is executed by employing a MLP network after the extraction of significant information from ECG signals. The work employs the utilisation of three training strategies, including Backpropagation (BackP), Lavenberg-Marquardt (LevM), and Bayesian Regularisation (BayR) [8–10]. The system's architecture is based on a specific algorithm that categorises the patient's condition as either normal or abnormal. The P peak, QRS wave, and T peak serve as input parameters for the MLP network, from which information regarding amplitude and duration is extracted. The MLP network in this case was activated using the Tansig and Purelin activation functions.

The subsequent content of the paper is structured in the following manner. Part II of this study entails a comprehensive overview of the research conducted on ANNs and MLP networks. Additionally, an elucidation of ECG data is presented. The explanation of the data sample can be found in Section III. Section V provides the concluding remarks after Section IV, which presents the findings and discusses specific points of contention.

II. Multilayer Perceptron Network

The construction of a computer network known as ANN involves the use of neural stem cells derived from the human brain, specifically neurons. Neurons, which are diminutive cellular entities, constitute the structural components of the human brain. ANNs are computational models that aim to simulate the structure and functionality of neurons in the human brain with the goal of replicating their ability to process information and exhibit creative behaviour. Mathematical modelling can be used to describe ANNs in non-parametric grading/regression, data/database classification, and nonlinear function calculation. ANN are employed to imitate the behaviour of biological models of human neurons. ANNs have the capability to be trained in order to generate precise prediction results. While ANNs are often considered a potential substitute for brain functionality, it is crucial for them to possess the capability to perform at a comparable level to that of the human brain.

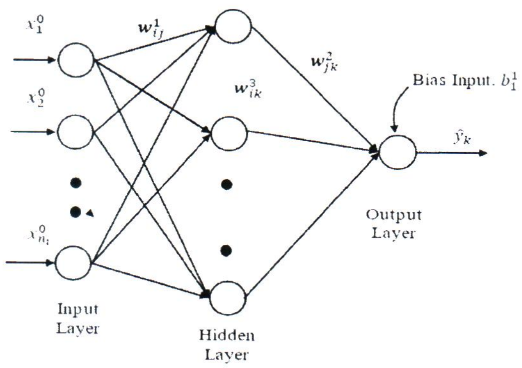

Fig. 1. The structure of the MLP network [13].

The created system may therefore function as a replacement for the original system. The MLP network is a highly prevalent application that serves a multitude of objectives [11–12].

Rosenblatt created the Perceptron, a model of an artificial neuron, in 1958 [7]. As illustrated in Figure 1, the individual proceeds to execute a series of network levels or sequences. ANNs are occasionally denoted as MLP networks in different domains of study. The input layer, hidden layer, and output layer of a MLP network is depicted in Fig. 1.

$$\sum_{j=1}^{n_h} w_{jk}^2\, \partial\left(\sum_{i=1}^{n_i} w_{ij}^1 x_i^0(t) + w_{k0}^1 b_j^1\right) \tag{1}$$

$for\ 1 \leq j \leq n_h$ and $1 \leq k \leq m$

The weight that connects the MLP network's input layer to the hidden layer itself is shown in Equation 1 as w_{ij}^1. The weights w_{jk}^2 and w_{jk}^2 reflect the connections between the hidden and output layers of the MLP network, respectively. In Equation 1, the input parameter, indicated as x_i^0, is feeding the MLP network through the input layer, whereas b_j^1 is the threshold of the hidden node. The MLP network, indicated as $\partial(\cdot)$ in Equation 1, has been activated in this work using the Tansig activation function. Equation (1) can be used to find the values of w_{ij}^1, w_{jk}^2 and b_j^1 based on a suitable technique and put at the minimal vector but allow for recurrence at each iteration.

The MLP network is trained using three distinct training procedures. The BackP, LevM, and BayR algorithms are three computational methods that have been developed and utilised in several fields of study. Bayesian Regularisation (BayR) employs a stochastic tabulation approach, whereas the training algorithms for Backpropagation (BackP) and Levenberg-Marquardt (LevM) rely on deterministic tabulation. A recent study investigated the performance of an MLP network using Tansig and Purelin activation functions. This study does a comparative analysis of various activation functions and training approaches. Tansig activation function is associated with a bipolar sigmoid. The output of the Tansig function range from -1 to + 1 but Purelin activation function operated within nominal parameters and behavior closely like linear function.

A. Bayesian Regularization (BayR) Algorithm

D. Bayes rule is a training/learning method that was developed by Thomas Bayes. Given the D data, the Bayes rule can be used to determine θ, also known as posterior probability [8]. Typically, the posterior probabilities produced cover the whole range of potential values for. Bayes rule could have come from Eq. 2.

$$p(\theta|D) = \frac{p(D|\theta)}{p(D)} \tag{2}$$

Following $p(\theta|D)$, which is referred to as the likelihood of the data before the probability of data, comes $p(\theta)$, which is the initial probability of a parameter. The probability distribution of the network weights was then modified by this general principle as it was applied to the MLP network. The weights of the MLP network are the w that it receives from the training data, $p(w|D)$. One can visualise the posterior distribution by

$$p(w|D) = \frac{p(D|w)p(D)}{p(D)} = \frac{p(D|w)}{\int p(D|w)\,p(w)dw} \tag{3}$$

Using the Bayesian equation as a guide, it can be shown that, as illustrated in Figure 3, the BayR training algorithm taught and optimised the weights that may modify our belief regarding the influences, from the initial $p(w)$ to the posterior $p(D|w)$.

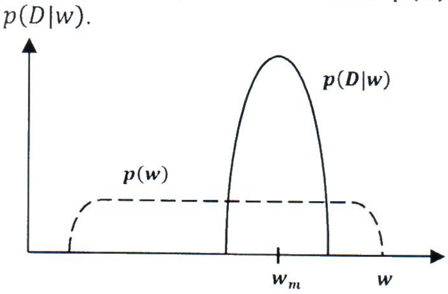

Fig. 2. Change pre weights to posterior/post weights [12].

B. Levenberg Marquardt (LevM) Algorithm

The functioning of the LevM training algorithm is based on a deterministic gradient tabulation. The LevM algorithm has made enhancements to the BackP algorithm. The profitability of the LevM method, following the training of the MLP network, is evaluated through a comparative analysis with the conventional BackP training technique. Furthermore, the LevM training method exhibits superior prediction performance compared to the classic BackP training algorithm, while also demonstrating a slightly faster convergence rate. The presence of relative stability persists [9]. The LevM training technique was developed to expedite second-order training and circumvent the computational challenges associated with the Hessian matrix, utilising a quasi-Newton approach. The calculation of the sum of squares is accomplished using the LevM function, and the resulting estimated Hessian matrix is presented below.

$$H = J^T J \tag{4}$$

and the gradient of the matrix is calculated as:

$$g = J^T \rho \tag{5}$$

where J is a Jacobian matrix that calculates network error using bias and weight of the network. The array can be calculated using conventional BackP methods, which are less difficult than the calculations involving the Hessian matrix [9]. The Newton-like equation estimator is used by the LevM algorithm to update the equations in the Hessian matrix as follows:

$$\Delta w = -[J^T J + \mu I]^{-1} J^T \rho \tag{6}$$

where Δw is the estimator's updated weight, which is determined by the μ parameter. To prevent the parameter reaching 0, the Hessian matrix updated to the Newton-like equation. Since it increases BackP's momentum, the Newton estimator based BackP training algorithm can deliver better results. With more momentum, the BackP training algorithm can find the lowest local minima and produce the least amount of error. At each step that is effective, the decreases (minimize error). However, when the step function develops, the continues to rise. At each cycle, the minimum error must, however, decrease [9].

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

C. Back Propagation (BackP) Algorithm

The back propagation training approach has been used in a wide variety of control systems and ANN applications. Electrical signals generated by brain activity and coupled to electrical activity or electrical signals in the heart are regularly read ECGs. On the hands, feet, and chest are some electrodes. The electrical activity and electrical impulses generated will be recorded, and this information will be used to train the BackP algorithm. Typically, the algorithm instructs the network using the steepest descent approach. The BackP method can use the learning phase as a parameter or protocol to determine the value of derivatives for practical weight adjustments [8]. According to, BackP has updated the weight between the j[th] neuron of the hidden layer and the i-layer directly.

$$w_{ji}(t) = w_{ji}(t-1) + \Delta w_{ji}(t) \tag{7}$$

and

$$b_j(t) = b_j(t-1) + \Delta b_j(t) \tag{8}$$

The updated weight, $\Delta w_{ji}(t)$ and $\Delta b_j(t)$ given by:

$$\Delta w_{ji}(t) = \eta_w \gamma_j(t) x_i(t) + \alpha_w \Delta w_{ji}(t-1) \tag{9}$$

and

$$b_j(t) = \eta_b \gamma_j(t) + \alpha_b \Delta b_j(t-1) \tag{10}$$

The subscripts w and b in Equations 7 and 8 respectively denote the weight and threshold of the steepest decent estimator. The constants α_w and α_b regulate the likelihood of the previous parameter to the present benchmark. The likelihood momentum must be carefully chosen. BackP won't be able to draw lessons from the past if the momentum is set too high. If it is adjusted too low, however, the BackP will continue to behave as it did previously. The BackP training algorithm's η_w and η_b learning rates are noted, whereas $\gamma_j(t)$ represents an error that occurs at the j[th] neuron. The error afterwards changed into an input and returned to weight. Equation 11 indicates error at the output node on the assumption that the activation function is functioned in a linear parameter:

$$\gamma(t) = y_k(t) - \hat{y}_k(t) \tag{11}$$

where $y_k(t)$ is the predicted output, and $\hat{y}_k(t)$ is the desired output. The production at the hidden layer is

$$y_j(t) = F^{\cdot}(x_i(t)) \sum_j \rho_j^k(t)\, w_{jk}^2(t-1) \tag{12}$$

where $F^{\cdot}(x_i(t))$ is derived from $F(x_i(t))$ by respecting to $x_i(t)$. As a steepest descent approach, the BackP algorithm suffers from a poor convergence speed or rate. The pursuit of a global minimum may lead to local minima. It responds similarly to user-selected parameters [8].

III. Data Sample

The ECG signal samples utilised in this investigation were sourced exclusively from the MIT-BIH Repository's collection of recordings. Three out of the six distinct characteristics that are derived from each signal sample include the amplitude of P, QRS, and T peaks, as well as the duration of P, QRS, and T waves [14-15]. The characteristics will function as input parameters for the recently established MLP network. The Multi-Layer Perceptron (MLP) network, constructed with six input parameters, was employed to establish the mapping between the input and output correlations of the system. The MLP will be utilised to establish a model that captures the relationship between the input variables and the corresponding output (target) values inside the dataset. The MLP network underwent training using a dataset including 1000 instances. Out of these, 800 instances were utilised for the purpose of optimising and generalising the networks, while the remaining 200 instances were employed to evaluate the performance of the MLP network.

IV. RESULT AND DISCUSSSION

The current investigation utilised Matlab software and implemented the precise methodological analysis outlined in prior research [16-18]. This article presents two discrete interpretations pertaining to the classification and forecasting of data. Initial research was undertaken to determine the most effective architecture for a MLP network. The MLP network configuration was constrained to a maximum of ten concealed nodes throughout the analysis phase. By conducting this research, the optimal number of iterations for the MLP network will be determined. The computation time will lengthen in proportion to the design's complexity, which is determined by variables including the selected structure, training algorithm, and activation function.

TABLE I. MLP OPTIMUM STRUCTURE

Training Algorithm	Num. of iteration
BayR using Tansig	425
BayR using Purelin	616
LevM using Tansig	32
LevM using Purelin	35
BackP using Tansig	19
BackP using Purelin	23

TABLE II. ACCURACY PERFORMANCE ANALYSIS

Training Algorithm	Accuracy Performance (%)		
	Training	Testing	Overall
BayR using Tansig	96.92	95.72	96.32
BayR using Purelin	93.14	93.24	93.19
LevM using Tansig	94.04	92.12	93.08
LevM using Purelin	89.89	89.56	89.73
BackP using Tansig	89.84	87.28	86.56
BackP using Purelin	86.23	85.67	85.80

Table I presents the results of the BackP, LevM, and BayR training algorithms utilised to train the MLP network with respect to accuracy. Conversely, each training algorithm is activated using the Logsig and Purelin activation functions. Once more, in the second analysis, the optimal structure is implemented, having been derived from the results of the initial study. ECG signals are classified into distinct categories based on the patient's physiological condition during the second analysis. Tabulated in Table I is the optimal MLP network architecture. Each BackP, LevM, and BayR training algorithm was evaluated with two distinct activation functions to determine the accuracy of the predictions provided. The performance of the prediction capability during the training and assessment phases is subsequently presented in Table II as the second analysis.

The BayR technique, with the Tansig activation function, can train the MLP network by generating networks with the highest achievable accuracy. However, this strategy necessitates the utilisation of 425 hidden nodes, as illustrated in Tables I and II. The findings indicate that BayR has the capability to effectively train MLP networks for constructing the foundational network design, as depicted in Table I. In contrast to the BayR training algorithm and the LevM training algorithm, the BackP construction of the simplest network but unable to perform high accuracy. Table II illustrates that the MLP network utilising the BayR training algorithm with Tansig activation function outperforms the other networks in terms of accuracy.

V. CONCLUSION

The objective of this research is to perform an examination of the ANN classification methodology utilised in cardiac anomaly monitoring and classification in order to determine whether the anomalies are typical or pathological in nature. The results of the study suggest that ANN exhibits a considerable degree of precision in their predictions. This is accomplished by providing MLP networks with the amplitude and duration characteristics extracted from ECG signals as input parameters. Additionally, the results suggest that the BayR algorithm can generate more precise prediction outcomes when compared to the LevM and BackP training methodologies while fabricating networks with the most extensive topology. Despite Purelin activation function, the Tansig activation function, which can activate the input parameter, was selected as the activation function for this result.

ACKNOWLEDGMENT

The authors fully acknowledged the National Defence University of Malaysia (NDUM) for the approved fund, making this important research viable and effective.

REFERENCES

[1] M. A. Quiroz-Juárez, O. Jiménez-Ramirez, R. Vázquez-Medina, E. Ryzhii, M. Ryzhii and J. L. Aragón, "Cardiac Conduction Model for Generating 12 Lead ECG Signals With Realistic Heart Rate Dynamics," in IEEE Transactions on NanoBioscience, vol. 17, no. 4, 2018, pp. 525-532.

[2] U. Satija, B. Ramkumar and M. S. Manikandan, "A New Automated Signal Quality-Aware ECG Beat Classification Method for Unsupervised ECG Diagnosis Environments," in IEEE Sensors Journal, vol. 19, no. 1, 2018, pp. 277-286.

[3] A. Walinjkar and J. Woods, "Personalized wearable systems for real-time ECG classification and healthcare interoperability: Real-time ECG classification and FHIR interoperability," 2017 Internet Technologies and Applications (ITA), Wrexham, 2017, pp. 9-14.

[4] U. Satija, B. Ramkumar and M. S. Manikandan, "Automated ECG Noise Detection and Classification System for Unsupervised Healthcare Monitoring," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 3, 2017, pp. 722-732.

[5] F. R. Hashim, J. J. Soraghan, L. Petropoulakis and N. G. N. Daud, "EMG cancellation from ECG signals using modified NLAVMS adaptive filters," 2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES), Kuala Lumpur, 2014, pp. 735-739.

[6] F. R. Hashim, J. Adnan, K. A. Ahmad, S. H. S. Ahmad Jamil and Y. Januar, "MLP Based Tan-Sigmoid Activation Function for Cardiac Activity Monitoring," MATEC Web Conf., 255, 2019.

[7] S. Haykin. Neural Network and Learning Machine Learning, rentice Hall, New Jersey, 2008.

[8] A. Talwar, G. Lokhande, R. Jain and S. Singh, "Estimation of aerodynamic parameters using Cascade Forward Back Propagation," 2017 2nd International Conference on Telecommunication and Networks (TEL-NET), Noida, 2017, pp. 1-6.

[9] Ü. Şentürk, İ. Yücedağ and K. Polat, "Cuff-less continuous blood pressure estimation from Electrocardiogram(ECG) and Photoplethysmography (PPG) signals with artificial neural network," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4.

[10] D. Yue, F. Xu, Z. Zhang and Y. Jin, "Quad-pol reconstruction with wishart-Bayesian regularization," 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, 2017, pp. 3933-3936.

[11] N. A. Mat Isa, F. R. Hashim, F. W. Mei, D. A. Ramli, W. M. Wan Omar and K. Z. Zamli, "Predicting quality of river's water based on algae composition using artificial neural network," Industrial Informatics, 2006 IEEE International Conference on, 2006, pp. 1340-1345.

[12] D. A. Ramli, J. M. Saleh, F. R. Hashim and N. A, Mat Isa, "Multilayered perceptron (MLP) network trained by recursive least squares algorithm," Computers, Communications, & Signal Processing with Special Track on Biomedical Engineering, 2005. CCSP 2005. 1st International Conference on, 2005, pp. 288-291.

[13] MIT-BIH Arrhythmia Database – PhysioNet cited from www.physionet.org/physiobank/database/mitdb

[14] J. Adnan, N. G. Nik Daud, S. Ahmad, M. H. Mat, M. T. Ishak, F. R. Hashim and M. M. Ibrahim, "Heart abnormality activity detection using multilayer perceptron (MLP) network," AIP Conference Proceedings, 1930(1), 2018, pp. 020013.

[15] P. Mitra, S. Mitra and S. K. Pal, "Staging of Cervical Cancer with Soft Computing," IEEE Transaction on Biomedical Engineering. vol. 47, no. 7, 2016, pp. 934-940.

[16] S. H. Syam Ahmad Jamil, J. A. Kadir, F. R. Hashim, B. Mustapha, N. S. Hasan and Y. Januar, "Optimization of ECG Peaks for Cardiac Abnormality Detection using Multilayer Perceptron," 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 2020, pp. 169-173.

[17] F. R. Hashim, S. H. Syam Ahmad Jamil, J. A. Kadir, N. S. Hasan, B. Mustapha and Y. Januar, "Tansig Based MLP Network Cardiac Abnormality," 2019 9th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 2019, pp. 199-203.

[17] A. C. Idris, M. R. Saad, Rahman, M. R. A. Rahman, F. R. Hashim and K. Kontis, " Experimental Validation of Artificial Neural Network (ANN) Model for Scramjet Inlet Monitoring and Control," International Journal of Recent technology and Engineering, 7(5), pp. 558-563.

[18] A. C. Idris, M. R. Saad, Rahman, M. R. A. Rahman, F. R. Hashim and K. Kontis, " Experimental Validation of Artificial Neural Network (ANN) Model for Scramjet Inlet Monitoring and Control," International Journal of Recent technology and Engineering, 7(5), pp. 558-563.